DUPLICATE ALSO

**FORECASTING SYSTEMS TECHNICAL NOTE NO. 5**

# COMPARATIVE VERIFICATION OF TROPICAL CYCLONE TRACK FORECASTS 1995: UKMO v. ECMWF

A.M.RADFORD

JULY 1996

ORGS UKMO F

National Meteorological Library
FitzRoy Road, Exeter, Devon. EX1 3PB

DUPLICATE ALSO

NATIONAL METEOROLOGICAL LIBRARY

**FORECASTING SYSTEMS TECHNICAL NOTE NO. 5**

# COMPARATIVE VERIFICATION OF TROPICAL CYCLONE TRACK FORECASTS 1995: UKMO v. ECMWF

A.M.RADFORD

JULY 1996

Forecasting Systems
Meteorological Office
London Road
Bracknell, Berkshire
RG12 2SZ

This paper has not been published. Permission to quote from it should be obtained from the Head of Forecasting Systems.

## 1. Introduction

Semi-automated verification of the Unified Global Model (UGM) forecasts of tropical cyclone (TC) tracks has been carried out for several years within Forecasting Systems. Until recently the verification has been in isolation, the only method of comparison being against CLIPER, which is a standard, 'no-skill' forecast based on climatology and persistence (Neumann, 1972).

External verification of the UGM forecasts has been carried out routinely for a few years by the Japanese Meteorological Agency (JMA). They compare TC track forecasts from their own model with those from the UGM and ECMWF[1]. Until recently the results have been less than encouraging for the UGM, with the ECMWF model being the most accurate of the three models. However the results for 1994 indicated that the UGM had the best performance of the 3 models (Figure 1). This improvement can be attributed to changes in the physical parameterisation schemes, implemented in January 1994, and the new TC initialisation scheme, introduced in October 1994 (Heming, Chan & Radford, 1995).

The fact that the JMA scores prior to 1994 indicated that the UGM errors were larger than ECMWF gave rise to concern since, unlike most other global modelling centres, ECMWF does not use a bogus scheme to initialise TCs in the model. It has even been suggested that bogus data can have a negative effect on the subsequent forecast, not only of TC tracks but also extra-tropical systems (Miller, *pers. comm.*). Although we have limited evidence to support isolated cases of this happening, e.g. Heming (1990), there is also overwhelming evidence in favour of bogussing (Heming, 1993).

Considering that the JMA verification was carried out on a relatively coarse grid, and was for the north-west Pacific area only, it was decided to perform our own comparative verification of the ECMWF forecasts. This paper describes the results of that study for the calendar year of 1995. Following a brief description of the verification method in section 2, the main results are presented in section 3, and concluding comments in section 4.

## 2. Verification method

In order to locate a TC within the model the 850hPa relative vorticity is used rather than, say, the low-level wind or mean sea-level pressure; this is in accordance with WMO recommendations. For the model analysis and 24-hour forecast, the interpolated maximum value of 850hPa relative vorticity within a search area centred on the observed position is considered to be the centre of the TC. For subsequent forecasts (T+48, T+72, T+96 and T+120) the search area centre is determined by linear extrapolation of the model's forecast movement over the preceding 24 hours. The search area radius used for forecasts is larger than that used for analyses to allow for deviation from a straight track. Full details of the verification scheme can be found in Heming (1994). In order that the ECMWF fields, when presented to the verification system, are as comparable as possible with the UGM fields, they have been extracted on a grid of 1.25° longitude by 1° latitude.

---

[1] The JMA verification is carried out using mean sea-level pressure on a 2.5° grid, for the north-west Pacific ocean basin only.

Several statistics are calculated routinely. These include mean errors in the zonal and meridional directions, mean errors resolved into along-track (AT) and cross-track (CT) components, as well as the more straightforward mean direct positional errors (DPE). [A complete definition of these errors is given in Appendix A.] A skill score is also computed using the CLIPER forecast as the standard.

As well as calculating mean statistics globally, or hemispherically, the statistics are also divided into six ocean basins - western North Pacific (NWP), eastern North Pacific (NEP), North Atlantic (NAT), North Indian (NI), South-West Indian (SWI) and Australian (AUS). This allows us to infer whether there are systematic differences in the model's performance in different areas of the globe.

Verification is carried out on two time-scales. Firstly, real-time verification statistics for a particular TC are calculated as soon as that TC has dissipated. The verifying observations used are the positions received in the advisory bulletins received over the GTS (e.g. from the National Hurricane Center, Miami). However, some time later the appropriate regional centre prepares 'best-track' information. In some cases this may not be until the end of the season. Additional observations that were not available at the time are used and anomalies in the track positions are removed, where appropriate, by interpolation. The statistics presented in this paper have been calculated with respect to best-track observations. Comparisons between real-time and best-track verification are now published routinely (e.g. Heming, 1996).

Statistics have been calculated for forecasts from 12 UTC only, since ECMWF only runs operational forecasts from this time. Analysis statistics have been calculated at both 00 and 12 UTC, so the sample sizes are proportionately larger.

## 3. Results

*a) Percentage Forecast*

Since we do not verify forecast runs that begin before the TC has developed (i.e. we do not verify TC genesis) the maximum possible sample size will naturally decrease with forecast lead time. One of the criteria necessary for the verification scheme to provide statistics is that the 850 hPa relative vorticity of the forecast TC centre must be above a critical value (see Heming, 1994). If the model dissipates the circulation too quickly, this will lead to a reduction in the sample of forecasts actually verified. The ratio of the actual sample size and the maximum possible is referred to as the 'percentage forecast'. Since the verification scheme forces verification of the analysis, regardless of the strength of circulation, the 'percentage analysed' is 100% for both models.

[Note that if the model retains the TC for too long - i.e. after it has dissipated in reality - there is no record of this in the statistics. Verification is only performed if an observation is present within 6 hours of the verifying time.]

Figure 2 shows the relationship between percentage forecast and forecast lead time for UGM and ECMWF. It is clear that the UGM has a tendency to dissipate TCs more quickly than ECMWF. However, it is interesting to see that this behaviour is not the same in all ocean basins. Looking at 72-hour forecasts in more detail, Figure 3 shows (for each basin) the actual number of forecasts verified compared with the maximum possible. For

five of the six basins, ECMWF maintains more TCs at T+72 than UGM. Indeed in the north Indian and Australian basins the percentage forecast is 100%, although the sample sizes are relatively small. The performances of the two models are comparable in the south-west Indian ocean, but ECMWF clearly forecasts a higher percentage of TCs in both basins of the north Pacific. However in the north Atlantic the UGM outperforms ECMWF in this respect. There is no obvious reason why the performance in the north Atlantic should be any different from the rest of the globe.

*b) Homogeneous Sample*

For the purposes of comparison, the statistics presented in the remainder of this paper have been calculated using a homogeneous sample. The sample sizes for each basin are given in Table 1.

| Basin | Analysis | T+24 | T+48 | T+72 | T+96 | T+120 |
|---|---|---|---|---|---|---|
| NWP | 237 | 93 | 70 | 46 | 32 | 23 |
| NEP | 101 | 40 | 30 | 20 | 13 | 9 |
| NAT | 245 | 97 | 83 | 65 | 53 | 42 |
| NI | 21 | 8 | 5 | 2 | 0 | 0 |
| SWI | 125 | 53 | 42 | 33 | 26 | 19 |
| AUS | 70 | 25 | 16 | 8 | 4 | 2 |
| Global | 799 | 316 | 246 | 174 | 128 | 95 |

Table 1: Homogeneous sample sizes used to calculate verification statistics

1. Mean DPE

The mean direct positional error (DPE) represents the average distance between the forecast (or analysed) centre of circulation and the best-track location.

Figure 4 shows the comparative rates of error growth for all ocean basins combined. Any advantage that ECMWF may have due to a longer data cut-off time (around 8 hours compared with 3 hours 20 minutes) is obviously outweighed by the bogus data present in the UGM analysis. However, although the UGM begins with an advantage at the analysis time (62 km *cf.* 181 km) this is gradually eroded during the forecast period, so that the mean errors of the two models are within 20 km of each other by T+120. This difference in error growth rates can also be seen in the independent verification (1994, NWP) from JMA in Figure 1. It is presumably a feature of the models' formulations rather than any difference in the analysis system.

The only basin that shows any significant deviation from this pattern is the NEP (see Figure 5). Here the UGM error levels off after T+48 and remains below 400 km, while the ECMWF error continues to grow, although it should be noted that the sample size in this basin is quite small. The ECMWF mean analysis error is largest in the NEP basin at 223 km, probably reflecting the poor data coverage in the area.

## 2. Skill

We measure the skill of the TC track forecast by comparing it with a statistical model called CLIPER - a mixture of CLImatology and PERsistence (Neumann, 1972). The skill score is defined as:

$$\frac{\text{CLIPER DPE - Model DPE}}{\text{CLIPER DPE}} \times 100\%$$

Positive skill indicates that the model forecast is better than CLIPER.

The CLIPER model is only valid out to T+72, so the global mean skill scores for T+24, 48 and 72 are shown in Figure 6. Because of its nature, CLIPER is hard to beat at short forecast ranges, so the UGM's skill score of 12% at T+24 is impressive, especially when compared with ECMWF's negative skill of -31%. In line with the mean DPE, ECMWF improves more quickly, from a lower baseline, than UGM so that the skill scores at T+72 are much closer.

The only basins in which the UGM showed negative skill were the NEP at T+24 and AUS at T+72, although the sample size for the latter was only one forecast. It is notoriously difficult to beat CLIPER in the NEP basin, since the TCs tend not to recurve polewards. It is therefore not surprising to have a negative skill score in the early stages of the forecast (Figure 7). ECMWF forecasts showed no skill over the first 3 days in the NEP basin.

## 3. Along-Track and Cross-Track Errors

Along-track (AT) and cross-track (CT) errors are calculated by resolving the error vector into components along, and perpendicular to, the observed TC track (see Appendix A). They give an indication of the bias of the model's track forecast. Specifically, positive AT errors indicate a fast bias, while positive CT errors indicate a tendency to move to the right of the track in the northern hemisphere (to the left in the southern hemisphere).

Figure 8 shows a polar plot of the global mean AT and CT errors for each model. Both models have a significant slow bias (negative mean AT error), which gradually increases during the forecast period, UGM being slightly worse than ECMWF from T+72 onwards. The mean CT error is relatively small, although ECMWF approaches -100 km at T+120.

Looking at the statistics for individual basins (not shown), the AT errors consistently indicate a slow bias for both models with one exception - in the NEP basin, ECMWF develops a fast bias from T+72 onwards. On the other hand, the sign of the mean CT error varies from area to area, although both models have the same tendency. In the NWP they have a right-of-track bias (UGM slightly worse), indicating premature or incorrect recurvature. On the other hand, in the SWI basin the right-of-track bias shown by both models (ECMWF slightly worse) indicates a reluctance to recurve the TCs. The CT errors in the other basins are generally small.

4. Case Studies

Two case studies have been chosen from the North Atlantic hurricane season. The first, the forecast of Hurricane Felix from 12 UTC 14 August 1995, demonstrates that a good initial location does not necessarily result in an accurate forecast. Figure 9 shows a comparison between the actual track and the two model forecasts. Both models initialised the centre of Felix very close to the reported position, but the two ensuing forecasts diverged markedly. ECMWF steered the hurricane west north-west into South Carolina, while UGM took it on a more northerly track. Note that after 3 days, the centre of the hurricane in UGM was north of 40°N so was not verified (see Heming, 1994). In reality, the hurricane looked as if it would hit the US coastline around the Cape Hatteras area, but then recurved sharply. This was actually one of the few relatively poor forecasts of Felix by UGM, although it still showed skill over CLIPER. For the whole lifetime of the storm, the mean skill score was over 40% for T+24, T+48 and T+72.

The second example is the forecast of Hurricane Luis from 12 UTC 2 September 1995 (Figure 10). Again, the two forecasts diverged considerably, but this time the error in the ECMWF initialisation (291 km) was much larger than in UGM (31 km). The steering in the ECMWF forecast took the track west south-west, so that the T+48 forecast was very accurate (an error of only 125 km). However, it continued to track into the Caribbean and then north-west into the Dominican Republic by T+120. On the other hand, the UGM forecast was extremely good, indicating the recurvature away from Puerto Rico. It is interesting that the UGM forecast exhibited a slightly *fast* bias, contrary to the average performance. This also applied to the ECMWF forecast, at least until T+72.

## 4. Summary

Objective verification has been carried out of forecast TC tracks in the UGM and ECMWF model for the calendar year of 1995. Using the criterion that a TC is present in the model if the 850 hPa relative vorticity is above a critical value, the evidence shows that ECMWF tends to retain TC circulations longer into the forecast period than UGM, although this is not the case in the north Atlantic. The higher resolution of the ECMWF model probably explains why TCs last longer, although it is difficult to find a reason why the north Atlantic should be different. This area is probably one of the least data-sparse of the six ocean basins, and one would expect this to result in better-than-average ECMWF initialisations. Although this is indeed the case in terms of mean analysis error, this is not translated into successful retention of TC circulations.

UGM has lower mean direct positional errors and shows higher skill than ECMWF. This is almost certainly due to the fact that TC locations are initialised in UGM using a bogus scheme whereas ECMWF relies on conventional observations. However, the rate of error growth in ECMWF is smaller than UGM so that, by the T+120 range, the two models have comparable errors. This pattern is generally constant around the globe, although UGM has much lower errors than ECMWF in the north-east Pacific basin. When considering the bias in

the track predictions, there appears to be no significant difference between the models. Both show a slow bias, which is slightly larger in UGM.

**Acknowledgements**

Thanks go to Julian Heming for carrying out the verification and providing the statistics. Best-track observations were provided by NHC (Miami), JTWC (Guam), Météo France (La Réunion) and the Bureau of Meteorology (Melbourne).

**References**

Heming, J.T. (1990). The over-strengthening of a jet forecast over the USA. *Central Forecasting Case Study No. 28.*

Heming, J.T. (1993). A trial of the impact of intervention in the global model. *Central Forecasting Technical Note No. 12.*

Heming, J.T. (1994). Keeping an eye on the hurricane. *nwp Gazette, Vol.1 No.2, pp 3-7.*

Heming, J.T. (1996). Summary of tropical cyclone activity and forecasts in the 1995 northern hemisphere season. *Forecasting Systems Technical Note No. 4.*

Heming, J.T., Chan, J.C.L. and Radford, A.M. (1995). A new scheme for the initialisation of tropical cyclones in the UK Meteorological Office global model. *Meteorological Applications Vol. 2, pp 171-184.*

Neumann, C. J. (1972). An alternative to the HURRAN tropical cyclone forecast system. *NOAA Tech. Memo., NWS SR-62, 24pp.*

**Figure Captions:**

Figure 1. JMA comparative TC track verification 1994 (north-west Pacific basin only) - UKMO v. ECMWF v. JMA.

Figure 2. Percentage frequency of TCs forecast by UGM and ECMWF (for all basins).

Figure 3. Number of TCs forecast by UGM and ECMWF at T+72 stratified by ocean basin.

Figure 4. Mean direct positional error of TC track forecasts by UGM and ECMWF (for all basins).

Figure 5. Mean direct positional error of TC track forecasts by UGM and ECMWF (north-east Pacific basin only).

Figure 6. Mean TC track forecast skill scores by UGM and ECMWF with respect to CLIPER (for all basins).

Figure 7. Mean TC track forecast skill scores by UGM and ECMWF with respect to CLIPER (north-east Pacific basin only).

Figure 8. Polar plot of mean along-track and cross-track errors of TC track forecasts by UGM and ECMWF (for all basins).

Figure 9. Comparison between the actual track and the UGM and ECMWF forecasts of Hurricane Felix from 12 UTC 14 August 1995.

Figure 10. Comparison between the actual track and the UGM and ECMWF forecasts of Hurricane Luis from 12 UTC 2 September 1995.

**Figure 1 - JMA model inter-comparison 1994**

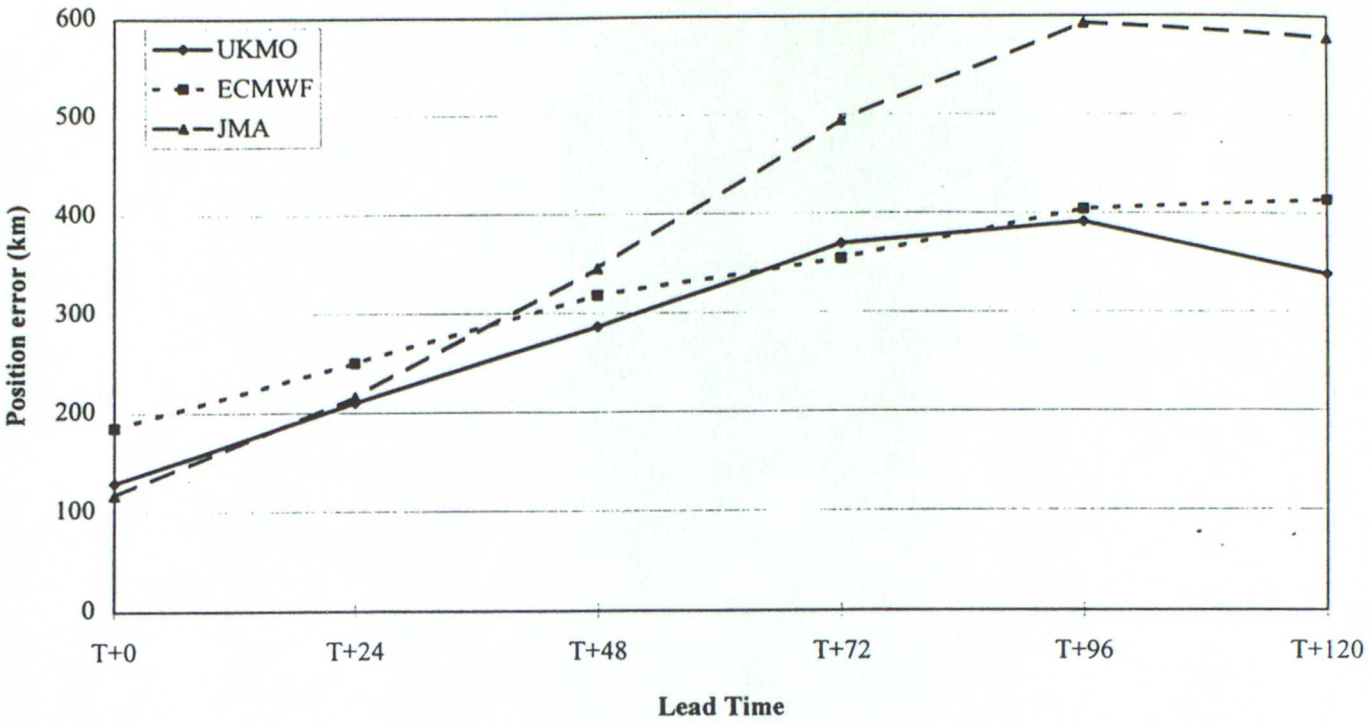



**Figure 2 - Percentage of TCs Forecast (all basins)**

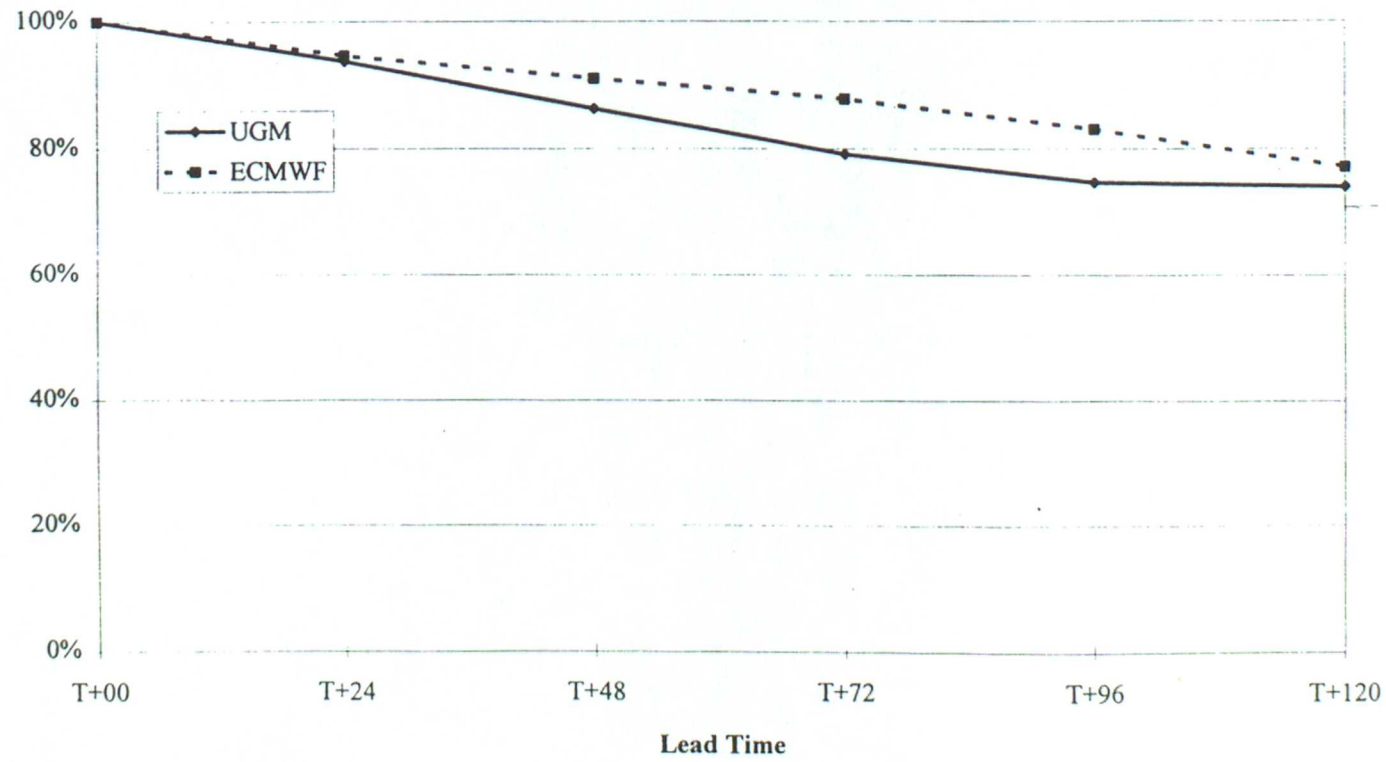


**Figure 3 - Frequency of TCs Forecast at T+72**

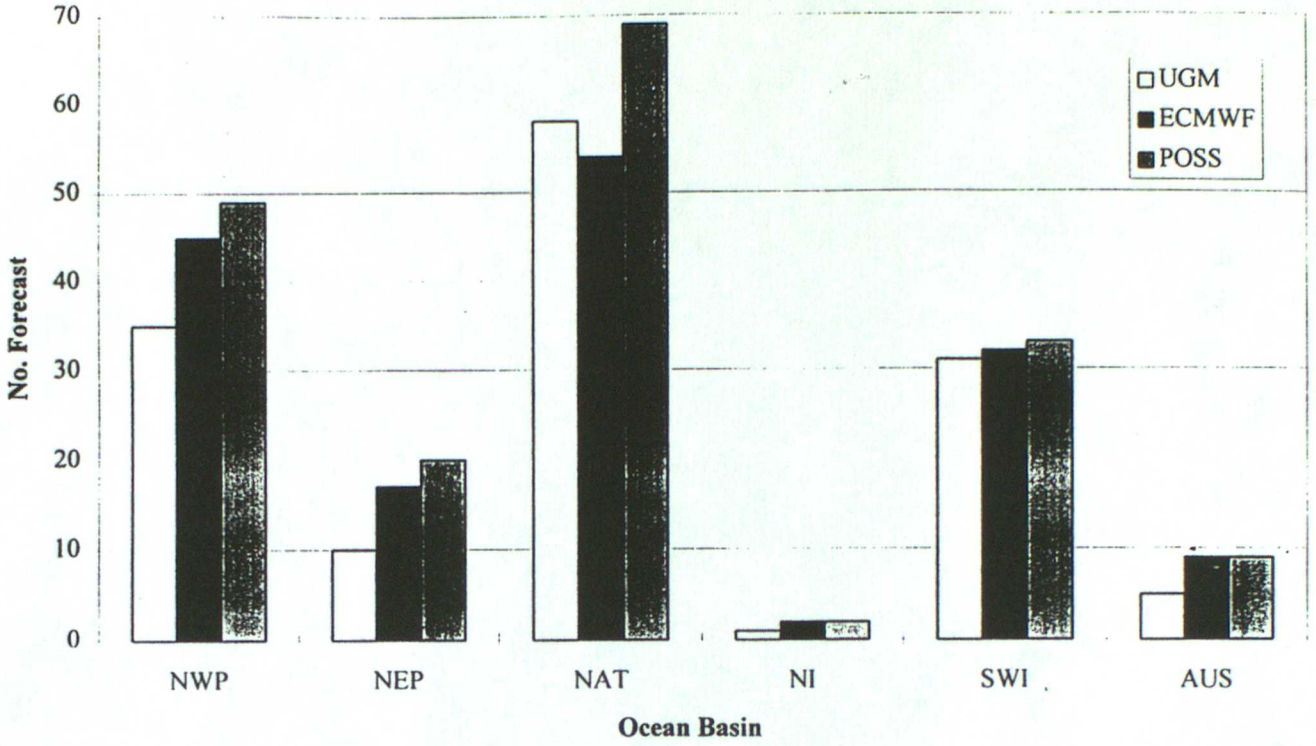



**Figure 4 - Mean Direct Positional Errors (all basins)**

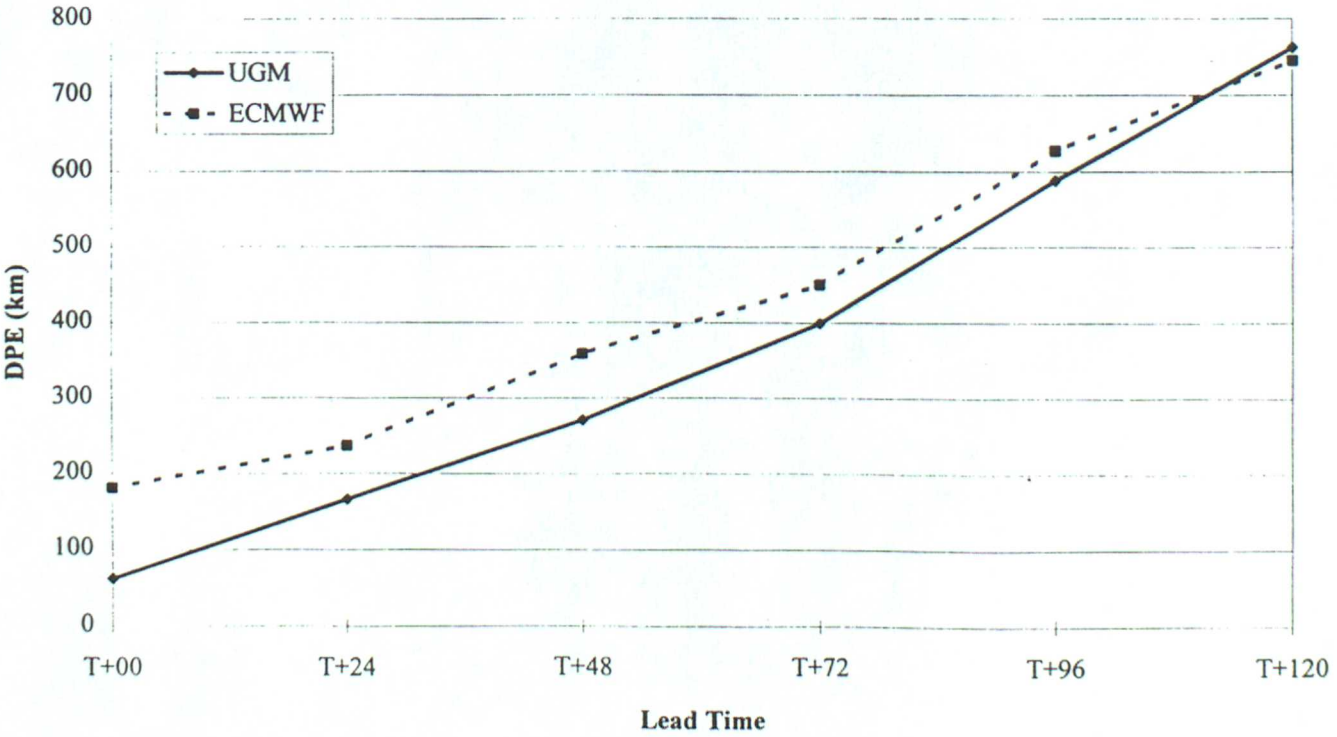


**Figure 5 - Mean Direct Positional Errors (NEP)**

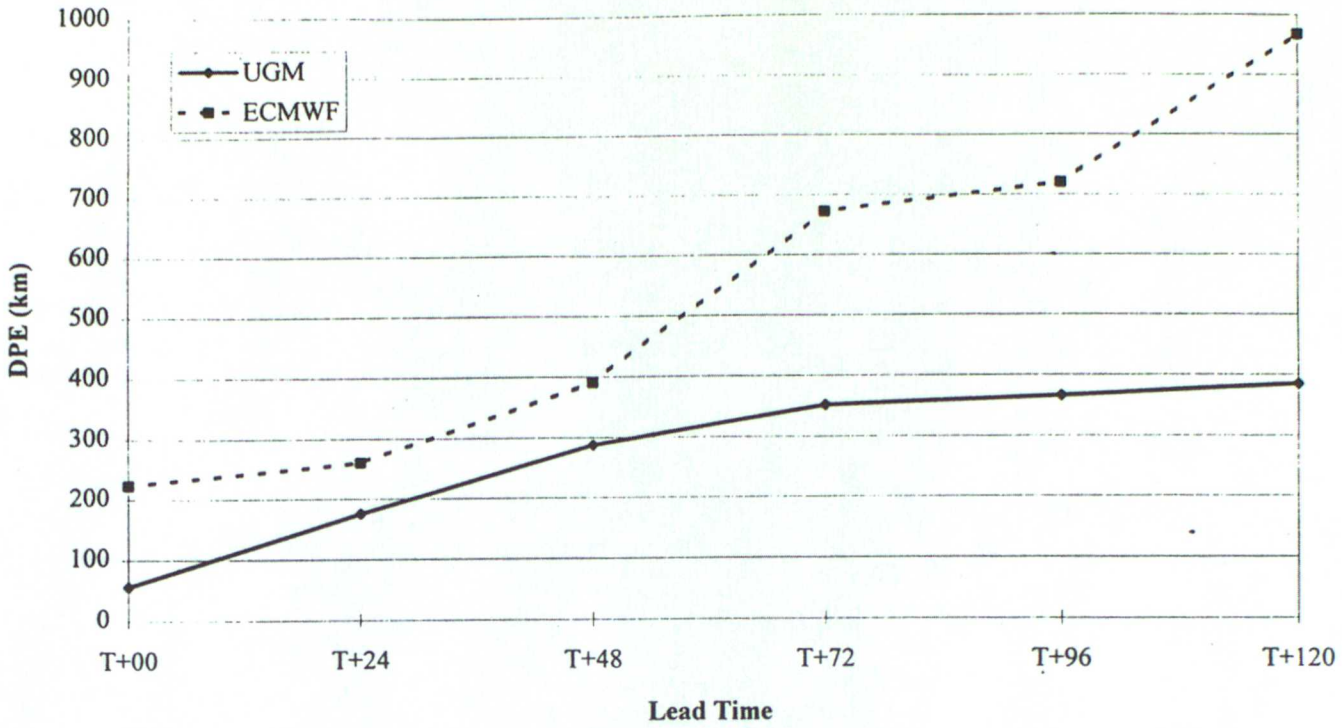



**Figure 6 - Mean Skill (all basins)**

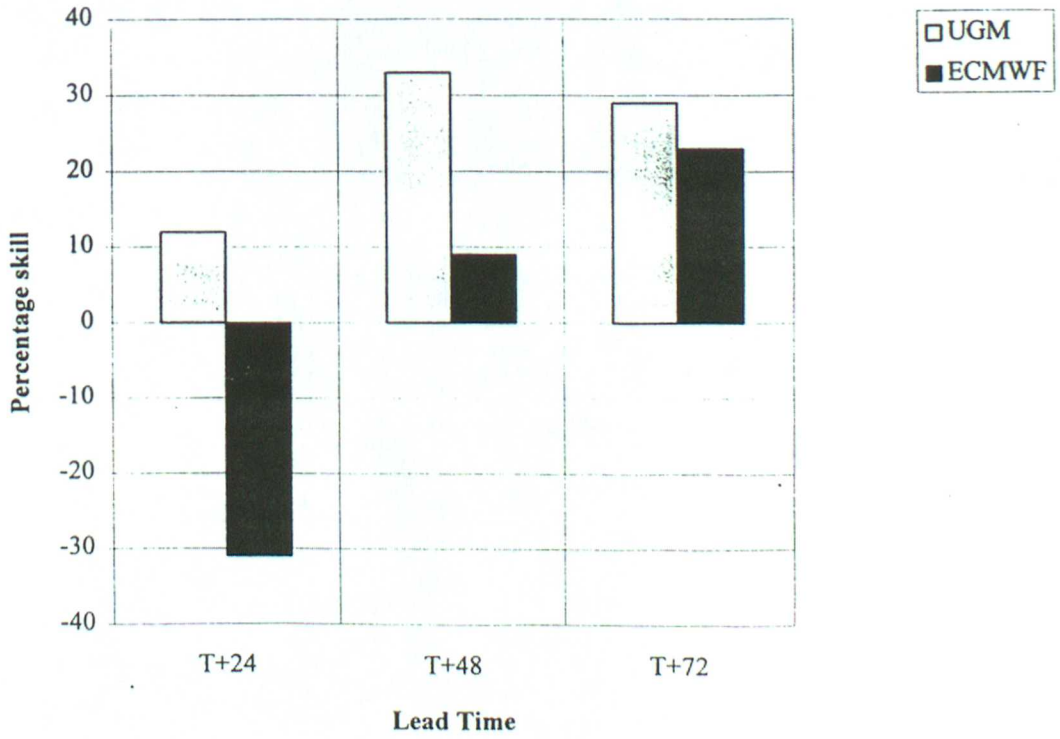


**Figure 7 - Mean Skill (NEP)**

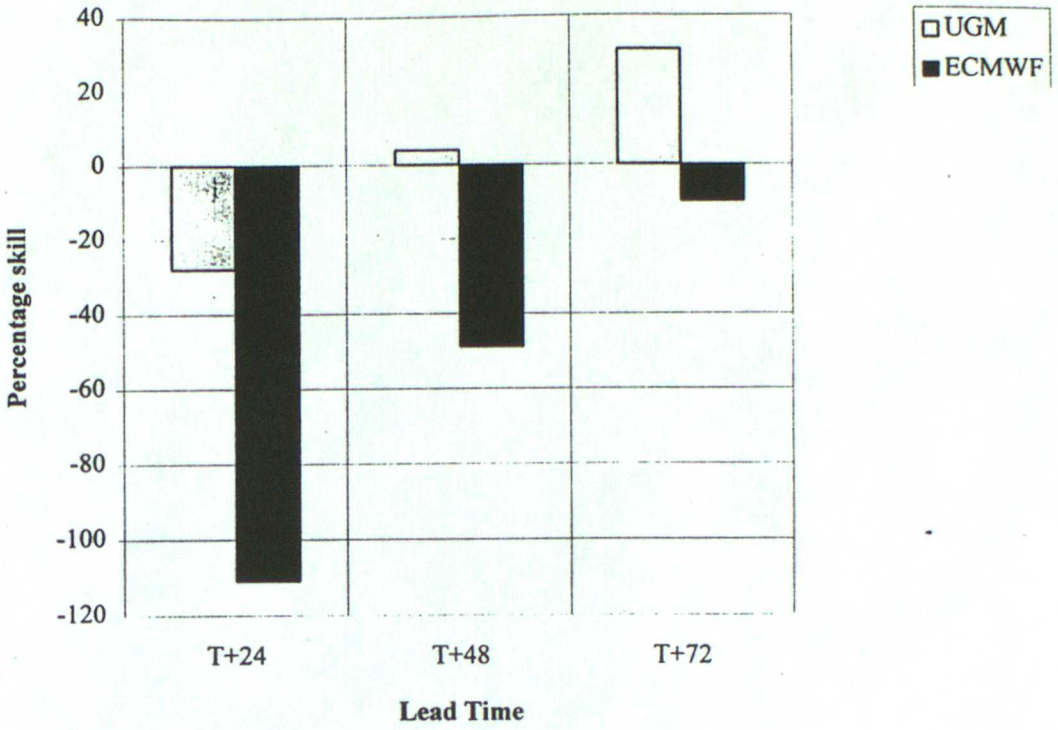



**Figure 8 - Mean Along- & Cross-Track Errors (all basins)**

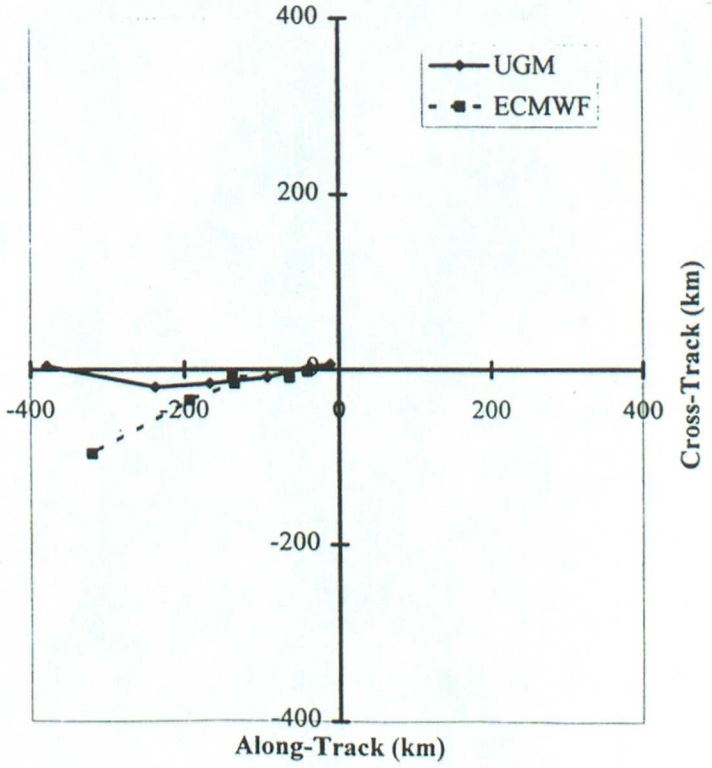
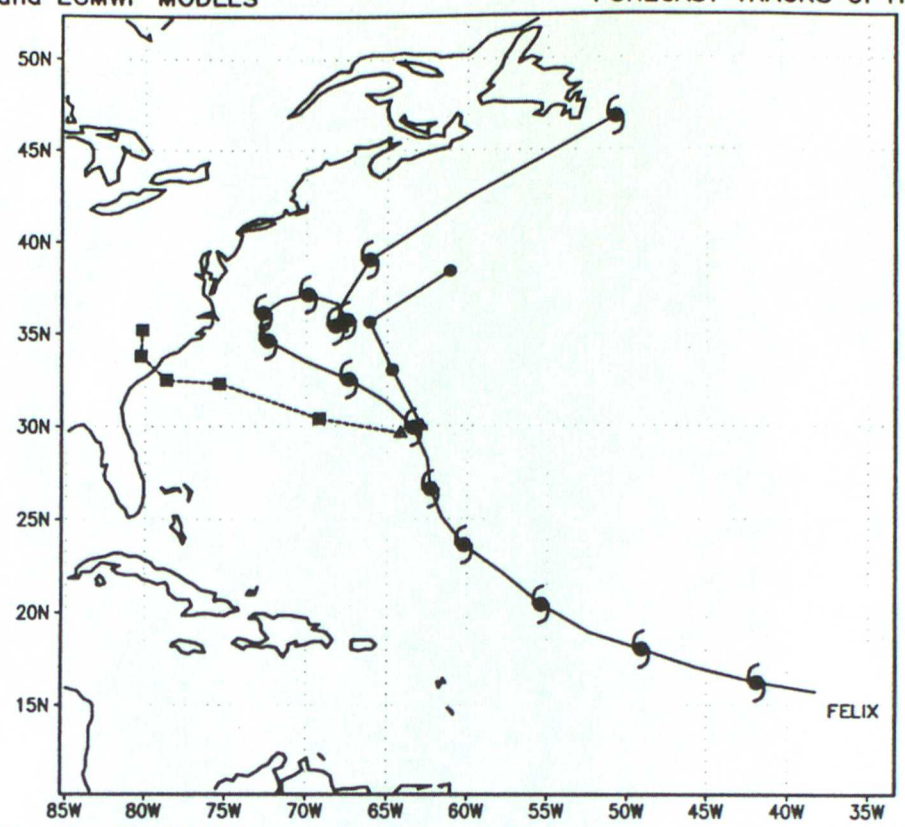


FIGURE 9

UKMO GLOBAL and ECMWF MODELS

FORECAST TRACKS of HURRICANE FELIX




UK 950814 EC 950814

KEY to FORECAST TRACKS

(Triangles denote analysed positions)

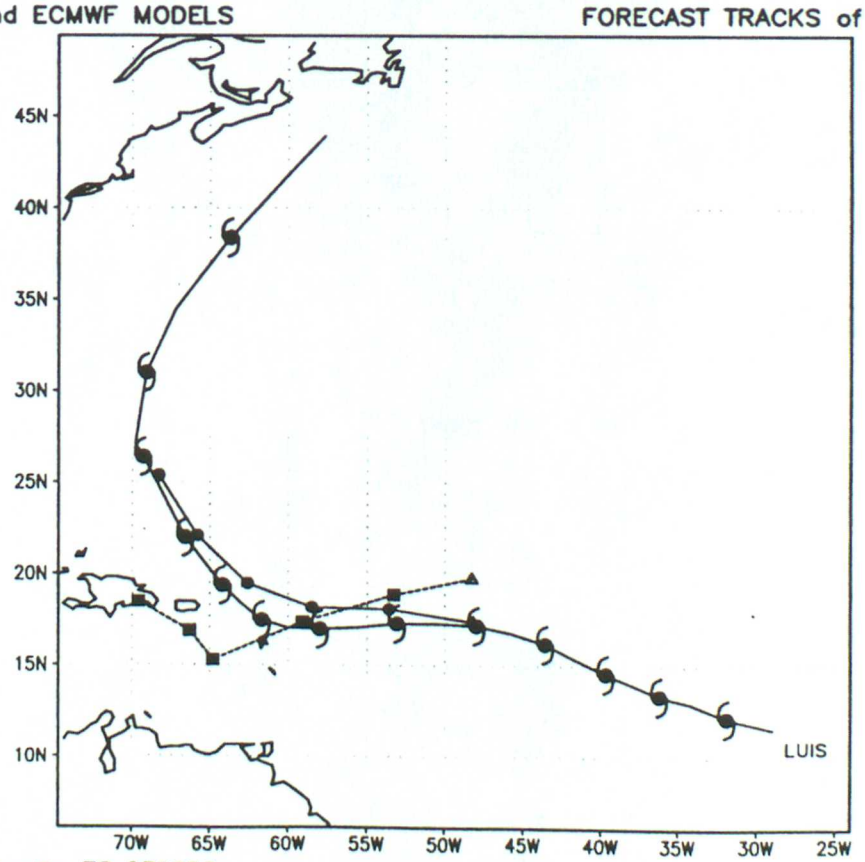
24 HOURLY BEST TRACK OBSERVED POSITIONS

DATE/TIME OF FIRST SYMBOL 12Z 09 AUGUST 1995

FIGURE 10

UKMO GLOBAL and ECMWF MODELS

FORECAST TRACKS of HURRICANE LUIS




UK 950902 EC 950902

KEY to FORECAST TRACKS

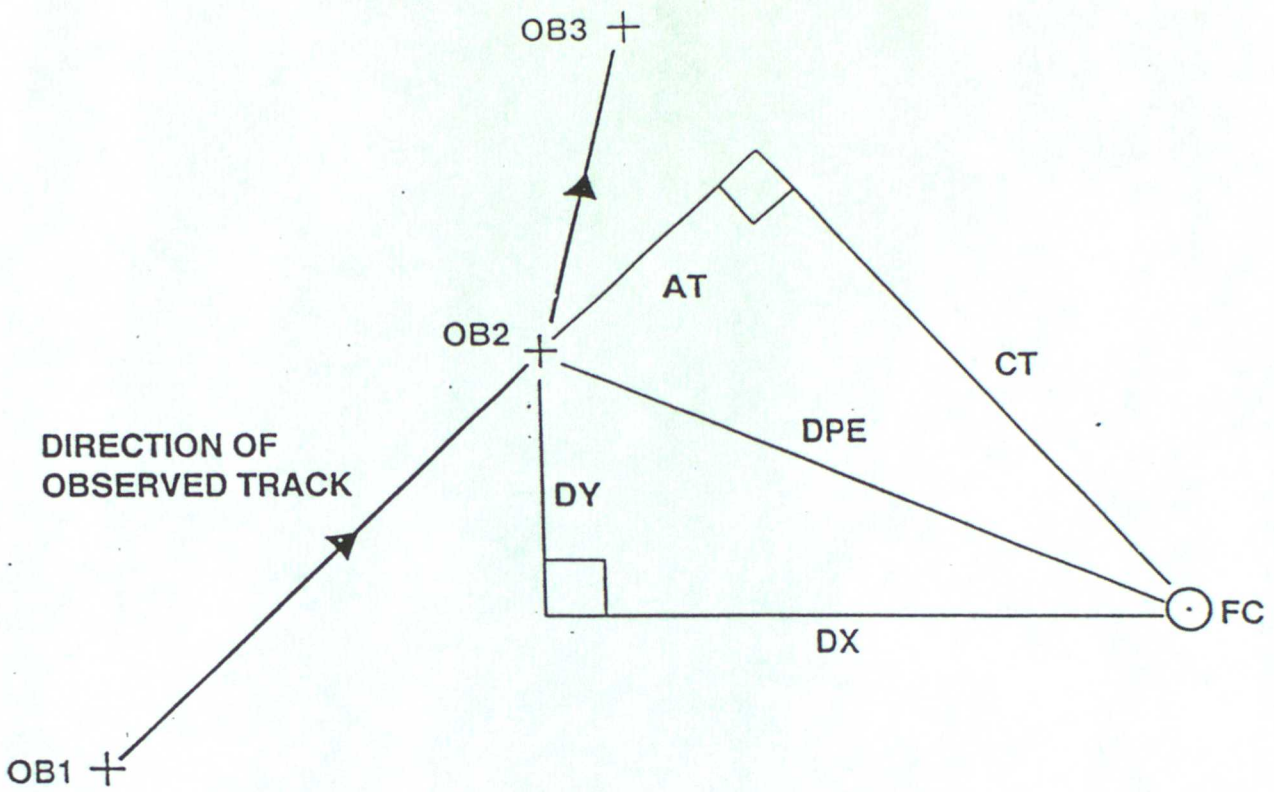
(Triangles denote analysed positions)

24 HOURLY BEST TRACK OBSERVED POSITIONS

DATE/TIME OF FIRST SYMBOL 12Z 29 AUGUST 1995

# APPENDIX A

## Diagrammatic Explanation of Forecast Errors




OB1-3 : Observed positions
FC : Forecast position verifying against observation OB2
DPE : Direct positional error
DX : Error in the East-West direction
DY : Error in the North-South direction
AT : Error in the Along Track direction
CT : Error in the Cross Track direction

All errors are measured in kilometres

## Sign Conventions

DPE values are always positive.
DX errors are positive if the forecast position lies eastwards of the observed position.
DY errors are positive if the forecast position lies polewards of the observed position.
AT errors are positive if the forecast position lies ahead of the observed position along the tropical cyclone track.
CT errors are positive if the forecast position lies right of the observed track in the northern hemisphere and left of the observed track in the southern hemisphere.